Lung Cancer (2005) **49**, 299—309

ELSEVIER

LUNG CANCER
IASLC

www.elsevier.com/locate/lungcan

Polymorphisms in folate metabolic genes and lung cancer risk in Xuan Wei, China

Min Shen[a,*], Nathaniel Rothman[a], Sonja I. Berndt[a], Xingzhou He[b], Meredith Yeager[c], Robert Welch[c], Stephen Chanock[c], Neil Caporaso[a], Qing Lan[a]

[a] *Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, MSC 7240, 6120 Executive Blvd., Bethesda, MD 20892-7240, USA*
[b] *Chinese Center for Disease Control and Prevention, Beijing, China*
[c] *CCR, National Cancer Institute, NIH, DHHS, Bethesda, MD 20892, USA*

Received 9 December 2004; received in revised form 31 March 2005; accepted 4 April 2005

KEYWORDS
Folate metabolism;
Single nucleotide polymorphism;
Lung cancer;
Polycyclic aromatic hydrocarbon;
Haplotype

Summary The aim of this study is to investigate the role of genetic polymorphisms in twelve folate metabolism genes on the risk of lung cancer in Xuan Wei, China, where the lung cancer mortality rate is among the highest and is mainly caused by indoor smoky coal emissions. A total of 122 incident primary lung cancer cases and 122 matched controls were enrolled. Three single nucleotide polymorphisms were associated with increased risk of lung cancer including homozygotes of the C allele of *CBS* Ala360Ala (OR: 4.02; 95% CI: 1.64—9.87), the 222Val allele of *MTHFR* (OR: 2.32; 95% CI: 1.34—4.03), and the C allele of *SLC19A1* Pro232Pro (OR: 1.83; 95% CI: 1.02—3.28). The distribution of *CBS* and *MTHFR* haplotypes differed between cases and controls ($P = 0.002$ and $P = 0.07$, respectively). In summary, three genetic variants in folate metabolism genes are associated with an increased risk of lung cancer in Xuan Wei, China.
© 2005 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Lung cancer is the leading cause of death from cancer worldwide with an estimated mortality of 31.43 for men and 9.53 for women per 100,000 in 2000 [1]. Tobacco smoking is the major attributable risk factor for the high prevalence of lung cancer across the world [2]. The lung cancer mortality rate in rural Xuan Wei County, Yunnan Province, is among the highest in China and is eight times the Chinese national average for women and four times the national average for men [3]. Although few women compared to men smoke in Xuan Wei, the mortality rates from lung cancer are similar between the sexes (27.7 and 25.3 per 100,000 for males and females, respectively). The extensive use of smoky coal indoors without adequate ventilation has been shown to cause the

* Corresponding author. Tel.: +1 301 4518791; fax: +1 301 4021819.
E-mail address: shenmi@mail.nih.gov (M. Shen).

0169-5002/$ — see front matter © 2005 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.lungcan.2005.04.002

high incidence of lung cancer in this population and accounts for more than 90% of lung cancer cases for both men and women [3,4]. Local smoky coal is a low-sulfur (0.2%), medium-volatile, bituminous coal and is used for cooking and heating in homes without chimneys [3]. During the burning of smoky coal, the indoor air concentration of particulate matter and extractable organic matter can reach as high as 24.4 and 17.6 mg/m^3, respectively [3], and the corresponding concentrations of benzo(*a*)pyrene, an indicator of PAHs, can reach as high as 19.3 $\mu g/m^3$ [5], which is comparable to exposure levels experienced by coke oven workers.

Folate, which is unsynthesizable by human cells, is considered to be a potential protective agent against cancer and is one of the components of fruits and vegetables [6]. Although not all epidemiologic studies have agreed [7,8], accumulating evidence suggests that reduced folate intake is a risk factor for lung cancer [9—12]. Folate, via its chemically reduced form tetrahydrofolate, is essential for the transfer of one-carbon units in the de novo synthesis of nucleotides. Reduced levels of N^5,N^{10}-methylenetetrahydrofolate lead to decreased synthesis of thymidylate from deoxyuridylate and consequently increase uracil misincorporation into DNA [13]. If not properly repaired, the misincorporation of uracil into DNA can cause DNA damage and chromosomal breaks [13]. In addition, folate is essential for maintaining normal DNA methylation patterns. Both global hypomethylation and hypermethylation of select CpG islands are thought to contribute to the pathogenesis of cancer [14]. Hypermethylation of CpG islands in the promoter regions of genes leads to transcriptional silencing, and hypermethylation of tumor suppressor genes and other genes involved in cell cycle control is thought to contribute to carcinogenesis. Several genes, including *CDKN2A* and *MGMT*, have been found to be hypermethylated in lung cancer tumors [15], suggesting that DNA methylation plays an important role in the pathogenesis of lung cancer. Through de novo DNA synthesis and methylation, folate is involved in DNA repair and low dietary folate intake was found to be associated with suboptimal DNA repair capacity [16].

Insufficient dietary folate intake is not the only reason for folate depletion. Tissue-specific folate concentration, varying distribution and aberrant function of co-enzymes in folate metabolism may play roles in maintaining the normal physical function of folate. Variants in at least two genes involved in folate metabolism have been shown to be associated with altered DNA methylation patterns [17]. In addition, a common polymorphism in the *MTHFR* gene, which converts N^5,N^{10}-methylenetetrahydrofolate to 5-methyltetrahydrofolate, has been found to lead to decreased enzymatic activity [18] and a shift in the tetrahydrofolate distribution [19]. Other genes involved in the metabolism of folate may also alter the distribution of folate and DNA methylation patterns.

Folate deficiency is a worldwide problem, especially in developing countries where people have less vitamin supplements and fortified food. It is a common but less realized health problem in the Chinese population [20—22], where the incidence of congenital neural tube defects is among the highest in the world [23,24]. Lack of adequate folate nutrition is thought to be a problem in Xuan Wei, because unfavorable economic conditions and lack of fresh fruits and vegetables make it difficult for people living in Xuan Wei to obtain good food sources of folate [4]. With reduced levels of folate, differences in folate metabolism could have a dramatic impact on the physiologic roles of folate in the body. Since genetic polymorphisms in folate metabolic genes may modify the ability of folate to protect against lung cancer, we studied the relationship between genetic polymorphisms of 12 folate metabolic genes (23 single nucleotide polymorphisms (SNPs)) and lung cancer risk in Xuan Wei, China.

2. Materials and methods

2.1. Study population

This was a population-based case-control study of lung cancer in Xuan Wei. Details of the study are described elsewhere [25]. A total of 122 newly diagnosed lung cancer cases and 1:1 individually matched controls were selected from March 1995 through March 1996. Matching conditions included sex, age (±2 years), village, and type of fuel currently used for cooking and home heating. The criteria for inclusion as a lung cancer case were positive histology or cytology results (105 cases, 86.1%) or clinically diagnosed cases who died within a 1-year period (17 cases, 13.9%). A standardized structured questionnaire was used to obtain information about demographic characteristics, life-time use of different types of coal, tobacco smoking, family history of lung cancer, and personal medical history.

2.2. Genotyping

DNA was extracted from sputum samples using phenol—chloroform extraction [26] and genotyped

by real-time PCR on an ABI 7900HT sequence detection system as described on the SNP500 website (http://www.snp500cancer.nci.nih.gov/) at Core Genotyping Facility of the National Cancer Institute. Of the 122 cases and 122 controls, DNA was successfully extracted from 119 cases and 113 controls and more than 95% of DNA samples were successfully genotyped for all candidate SNPs.

2.3. Statistical analysis

An ever-smoker was defined as person who smoked at least one cigarette per day for 6 months or longer. The cutoff points for smoky coal use (t) and tobacco smoking (pack-year) were estimated based on the distribution of lifetime cumulative use of each in the controls. The Hardy-Weinberg equilibrium for each SNP was tested with Pearson χ^2 or exact test if any of the cell counts were small. Genotype data were analyzed with the homozygote of the common allele as the reference group. Because genotype data were not available for all cases and controls, unconditional logistic regression was used to estimate the odds ratios (ORs) and 95% confidence intervals (CIs) for the association between lung cancer risk and each SNP, adjusted for age, sex, and current fuel type. The results were similar when conditional logistic regression was used to analyze the data. Gene—environment interactions were tested on a multiplicative scale by adding product terms into a logistic regression model. False Discovery Rate (FDR)-adjusted *P* values were calculated using the Benjamini-Hochberg method [27] to assess if the obtained *P* values are still significant after multiple comparisons were taken into consideration.

Measures of pairwise linkage disequilibrium (LD) between SNPs within one gene were estimated using the program, HaploView (http://www.broad.mit.edu/personal/jcbarret/haploview/). Haplotype block structure was examined for SNPs within the same gene using the four gamete rule. Haplotypes were estimated for SNPs within the same haplotype block using the expectation-maximization algorithm, and overall differences in the frequency distributions of the haplotypes between cases and controls were tested using the permutation omnibus test available in SAS/Genetics. Individual haplotypes were also estimated with SAS/Genetics, and the effects of each haplotype were estimated using the best haplotype pairs in an unconditional logistic regression model with the most common haplotype as the reference. Data were analyzed with the Statistical Analysis Software, version 8.02 (SAS Institute Inc., 1996) if not specified elsewhere.

3. Results

Demographic features, including age, sex, ethnicity, education level, household income, dwelling type, and type of fuel source, were comparable between 122 case and 122 controls (Table 1). Smoky coal use but not smoking was associated with an increased risk of lung cancer, which is consistent with previous studies in Xuan Wei [28]. Twenty-three SNPs in twelve genes were genotyped in the study population. With the exception of *GGH* −353G > T ($P = 0.01$), *FPGS* Ex15 −260 C > T ($P = 0.03$), *MTHFD2* IVS1 +3323 T > C ($P = 0.01$), and *MTRR* His622Tyr ($P = 0.02$), the genotype frequencies for all other SNPs in the controls were consistent with Hardy-Weinberg equilibrium. The genotype frequencies for cases and controls and estimated associations with lung cancer risk are shown in Table 2. Three SNPs in three different genes displayed significant associations with lung cancer risk. Homozygous carriers of the C allele of *CBS* Ala360Ala were found to have a four-fold risk of lung cancer (OR: 4.02; 95% CI: 1.64—9.87), and variant carriers of either the 222Val of *MTHFR* (OR: 2.32; 95% CI: 1.34—4.03) or the C allele of *SLC19A1* Pro232Pro (OR: 1.83; 95% CI: 1.02—3.28) displayed an approximate two-fold risk of lung cancer. After adjusting for FDR, *P* values for *CBS* Ala360Ala and *MTHFR* Ala222Val remained statistically significant ($P < 0.05$). The other variants evaluated in this study were not associated with lung cancer risk.

Table 1 Distribution of demographic features in lung cancer cases and controls[a]

	Cases (%) $N = 122$	Controls (%) $N = 122$	*P*-value[b]
Age			
<55	52 (43)	51 (42)	
≥55	70 (57)	71 (58)	0.90
Sex			
Male	79 (65)	79 (65)	
Female	43 (35)	43 (35)	1.00
Smoking[c]			
No	9 (11)	10 (13)	
Yes	70 (89)	69 (87)	0.81
Smoky coal use			
<130	51 (42)	72 (59)	
≥130	71 (58)	50 (41)	0.007

[a] Demographic data were previously reported [25] and DNA were extracted from 119 cases and 113 controls.
[b] *P*-value based on Pearson χ^2-test.
[c] Males only.

Table 2 Main effect of genetic polymorphisms in folate metabolism genes on lung cancer risk in Xuan Wei

Gene SNP (dbSNP ID)	Cases (%) *N* = 119	Controls (%) *N* = 113	OR[a]	95% CI	*P*-value	OR[b]	95% CI	*P*-value
BHMT								
IVS4 +52 C > T (rs567754)								
CC	37 (32)	27 (24)	Ref.			Ref.		
CT	58 (50)	64 (58)	0.66	0.36—1.22	0.18	0.57	0.30—1.08	0.08
TT	20 (17)	19 (17)	0.76	0.34—1.70	0.51	0.61	0.26—1.40	0.24
CT + TT	78 (68)	83 (76)	0.68	0.38—1.23	0.20	0.58	0.32—1.07	0.08
Trend					0.39			0.17
Ex8 +453 A > T (rs585800)								
AA	91 (77)	94 (83)	Ref.			Ref.		
AT	26 (22)	18 (16)	1.50	0.77—2.94	0.23	1.59	0.80—3.17	0.19
TT	1 (1)	1 (1)						
AT + TT	27 (23)	19 (17)	1.48	0.76—2.86	0.24	1.55	0.79—3.06	0.20
Trend					0.28			0.23
CBS								
IVS3 −1489 C > A (rs397589)								
CC	80 (70)	76 (68)	Ref.			Ref.		
CA	29 (25)	33 (29)	0.86	0.46—1.51	0.56	0.83	0.45—1.53	0.55
AA	6 (5)	3 (3)	1.90[c]	0.39—12.11	0.50			
CA + AA	35 (30)	36 (32)	0.93	0.53—1.63	0.79	0.93	0.52—1.67	0.82
Trend					0.91			0.84
Tyr233Tyr								
Ex8 +33 C > T (rs234706)								
CC	111 (94)	103 (91)	Ref.			Ref.		
CT	7 (6)	10 (9)	0.65	0.24—1.78	0.40	0.63	0.23—1.74	0.37
Ala360Ala								
Ex12 +41 T > C (rs1801181)								
TT	30 (26)	43 (38)	Ref.			Ref.		
TC	62 (53)	60 (54)	1.49	0.83—2.68	0.18	1.65	0.90—3.02	0.11
CC	25 (21)	9 (8)	4.02	1.64—9.87	0.002	4.34	1.73—10.86	0.002
TC + CC	87 (74)	69 (62)	1.81	1.03—3.19	0.04	2.01	1.12—3.60	0.02
Trend					0.003			0.002
FPGS								
Ex15 −260 C > T (rs10106)								
CC	52 (45)	49 (44)	Ref.			Ref.		
CT	49 (43)	56 (51)	0.82	0.47—1.42	0.47	0.71	0.40—1.26	0.24
TT	14 (12)	5 (4)	2.62	0.88—7.87	0.08	2.35	0.77—7.20	0.13
CT + TT	63 (55)	61 (56)	0.96	0.57—1.64	0.89	0.84	0.49—1.46	0.54
Trend					0.43			0.71
FTHFD								
Leu395Leu								
Ex10 −40 G > T (rs2305230)								
GG	80 (68)	65 (63)	Ref.			Ref.		
GT	29 (25)	31 (30)	0.74	0.40—1.37	0.34	0.79	0.42—1.48	0.46
TT	8 (7)	7 (7)	0.93	0.32—2.74	0.90	1.04	0.34—3.20	0.95
GT + TT	37 (32)	38 (37)	0.78	0.44—1.37	0.38	0.83	0.46—1.50	0.54
Trend					0.51			0.69
Asp793Gly								
Ex21 +31 A > G (rs1127717)								
AA (Asp/Asp)	91 (77)	88 (79)	Ref.			Ref.		
AG (Asp/Gly)	25 (21)	21 (19)	1.15	0.60—2.21	0.67	1.15	0.58—2.25	0.69
GG (Gly/Gly)	2 (2)	2 (2)	0.97[c]	0.07—13.61	1.00			
AG (Asp/Gly) + GG (Gly/Gly)	27 (23)	23 (21)	1.14	0.60—2.13	0.69	1.11	0.58—2.14	0.75
Trend					0.74			0.83
GGH								
−353 G > T (rs719235)								
GG	90 (76)	91 (81)	Ref.			Ref.		
GT	28 (24)	17 (15)	1.67	0.85—3.27	0.13	1.58	0.80—3.14	0.19
TT	0	4 (4)						
GT + TT	28 (24)	21 (19)	1.36	0.72—2.58	0.35	1.28	0.67—2.46	0.46
IVS7 −3001 C > T (rs1031552)								
CC	46 (40)	37 (33)	Ref.			Ref.		
CT	54 (46)	57 (51)	0.76	0.42—1.35	0.34	0.72	0.40—1.29	0.27
TT	16 (14)	17 (15)	0.75	0.34—1.70	0.50	0.73	0.32—1.68	0.46
CT + TT	70 (60)	74 (67)	0.75	0.44—1.31	0.32	0.72	0.41—1.26	0.25
Trend					0.38			0.33

Table 2 (*Continued*)

Gene SNP (dbSNP ID)	Cases (%) *N* = 119	Controls (%) *N* = 113	OR[a]	95% CI	*P*-value	OR[b]	95% CI	*P*-value
MTHFD2								
IVS1 +3323 T > C (rs1667627)								
TT	48 (42)	49 (45)	Ref.			Ref.		
TC	54 (47)	39 (36)	1.42	0.79–2.55	0.24	1.41	0.77–2.56	0.26
CC	13 (11)	21 (19)	0.63	0.28–1.42	0.26	0.64	0.28–1.45	0.28
TC + CC	67 (58)	60 (55)	1.14	0.67–1.96	0.63	1.13	0.65–1.97	0.65
Trend					0.61			0.62
MTHFR								
Ala222Val								
Ex4 +79 C > T (rs1801133)								
CC (Ala/Ala)	33 (28)	53 (48)	Ref.			Ref.		
CT (Ala/Val)	65 (56)	42 (38)	2.52	1.40–4.53	0.002	2.61	1.42–4.78	0.002
TT (Val/Val)	18 (16)	16 (14)	1.81	0.81–4.05	0.15	2.17	0.94–5.01	0.07
CT (Ala/Val) + TT (Val/Val)	83 (72)	58 (52)	2.32	1.34–4.03	0.003	2.49	1.41–4.42	0.002
Trend					0.03			0.01
Ala429Glu								
Ex7 −62 A > C (rs1801131)								
AA (Glu/Glu)	71 (62)	69 (63)	Ref.			Ref.		
AC (Glu/Ala)	41 (36)	34 (31)	1.18	0.67–2.08	0.57	1.07	0.60–1.91	0.82
CC (Ala/Ala)	2 (2)	6 (6)	0.32[c]	0.03–1.90	0.28			
AC (Glu/Ala) + CC (Ala/Ala)	43 (38)	40 (37)	1.04	0.60–1.80	0.88	0.95	0.54–1.66	0.84
Trend					0.71			0.49
MTHFS								
IVS2 −1411 T > G (rs622506)								
TT	30 (26)	36 (32)	Ref.			Ref.		
TG	56 (49)	55 (50)	1.21	0.65–2.24	0.54	1.20	0.64–2.26	0.56
GG	29 (25)	20 (18)	1.74	0.82–3.69	0.15	1.65	0.77–3.56	0.20
TG + GG	85 (74)	75 (68)	1.35	0.76–2.41	0.30	1.33	0.74–2.40	0.35
Trend					0.15			0.21
MTRR								
Ile49Met								
Ex2 −64 A > G (rs1801394)								
AA (Ile/Ile)	69 (58)	70 (62)	Ref.			Ref.		
AG (Ile/Met)	42 (36)	41 (36)	1.03	0.60–1.78	0.90	1.09	0.62–1.91	0.75
GG (Met/Met)	7 (6)	2 (2)	3.55[c]	0.64–35.92	0.17			
AG (Ile/Met) + GG (Met/Met)	49 (42)	43 (38)	1.15	0.68–1.95	0.60	1.22	0.71–2.11	0.47
Trend					0.32			0.23
Leu175Ser								
Ex5 +123 C > T (rs1532268)								
CC (Ser/Ser)	91 (78)	81 (74)	Ref.			Ref.		
CT (Ser/Leu)	24 (20)	25 (23)	0.85	0.45–1.61	0.62	0.91	0.47–1.76	0.79
TT (Leu/Leu)	2 (2)	4 (4)	0.44[c]	0.04–3.21	0.43			
CT (Ser/Leu) + TT (Leu/Leu)	26 (22)	29 (26)	0.80	0.43–1.47	0.47	0.86	0.46–1.62	0.65
Trend					0.36			0.53
Arg415Cys								
Ex9 −85 C > T (rs2287780)								
CC (Arg/Arg)	78 (68)	69 (63)	Ref.			Ref.		
CT (Arg/Cys)	32 (28)	38 (34)	0.75	0.42–1.33	0.32	0.68	0.38–1.22	0.20
TT (*Cyc*/Cys)	5 (4)	3 (3)	1.47[c]	0.27–9.82	0.72			
CT (Arg/Cys) + TT (Cyc/Cys)	37 (32)	41 (37)	0.80	0.46–1.39	0.43	0.74	0.42–1.30	0.30
Trend					0.66			0.52
His622Tyr								
Ex14 +14 C > T (rs10380)								
CC (His/His)	81 (72)	78 (71)	Ref.			Ref.		
CT (His/Tyr)	32 (28)	25 (23)	1.23	0.67–2.26	0.51	1.13	0.60–2.12	0.71
TT (Tyr/Tyr)	0	7 (6)						
CT (His/Tyr) + TT (Tyr/Tyr)	32 (28)	32 (29)	0.96	0.53–1.72	0.88	0.88	0.48–1.61	0.68
SHMT1								
Ex12 +236 C > T (rs1979276)								
CC	98 (83)	98 (88)	Ref.			Ref.		
CT	20 (17)	13 (12)	1.54	0.72–3.28	0.26	1.50	0.70–3.26	0.30
TT	0	1 (1)						
CT + TT	20 (17)	14 (12)	1.44	0.68–3.01	0.34	1.39	0.65–2.97	0.40

Table 2 (*Continued*)

Gene SNP (dbSNP ID)	Cases (%) *N* = 119	Controls (%) *N* = 113	OR[a]	95% CI	*P*-value	OR[b]	95% CI	*P*-value
SLC19A1								
Pro232Pro								
Ex4 −254 T > C (rs12659)								
TT	26 (22)	38 (34)	Ref.			Ref.		
TC	58 (49)	51 (46)	1.67	0.89−3.13	0.11	1.44	0.75−2.75	0.27
CC	34 (29)	23 (20)	2.17	1.05−4.51	0.04	2.12	1.00−4.49	0.05
TC + CC	92 (78)	74 (66)	1.83	1.02−3.28	0.04	1.65	0.90−3.02	0.11
Trend					0.04			0.05
Ex7 −233 G > T (rs1051296)								
GG	33 (28)	34 (31)	Ref.			Ref.		
GT	51 (43)	55 (51)	0.95	0.51−1.76	0.87	0.81	0.43−1.54	0.52
TT	34 (29)	19 (18)	1.86	0.88−3.92	0.10	1.85	0.86−3.98	0.12
GT + TT	85 (72)	74 (69)	1.18	0.66−2.10	0.57	1.07	0.59−1.93	0.83
Trend					0.13			0.16
TYMS								
IVS1 −405 C > T (rs502396)								
CC	49 (43)	43 (39)	Ref.			Ref.		
CT	53 (46)	52 (47)	0.89	0.51−1.56	0.68	0.84	0.48−1.50	0.57
TT	13 (11)	16 (14)	0.72	0.31−1.66	0.44	0.69	0.29−1.63	0.39
CT + TT	66 (57)	68 (61)	0.85	0.50−1.44	0.54	0.81	0.47−1.40	0.44
Trend					0.44			0.37
Ex7 +157 T > C (rs699517)								
TT	49 (42)	44 (40)	Ref.			Ref.		
TC	54 (46)	48 (43)	1.00	0.57−1.77	0.99	0.96	0.54−1.71	0.89
CC	15 (13)	19 (17)	0.70	0.32−1.55	0.38	0.68	0.30−1.53	0.35
TC + CC	69 (58)	67 (60)	0.92	0.54−1.56	0.75	0.88	0.51−1.52	0.64
Trend					0.48			0.42

[a] Adjusted for age, sex, and current fuel type by unconditional logistic regression.
[b] Adjusted for age, sex, current fuel type, pack-year of smoking, and coal use by unconditional logistic regression.
[c] Fisher's exact estimate and test for the parameter without adjustment for other factors.

In stratified analysis by smoky coal use, the effect of CC genotype of *CBS* Ala360Ala was stronger among light smoky coal users, and the effect of the C allele of *SLC19A1* Pro232Pro was restricted to light smoky coal users only (Table 3). The statistical test for multiplicative interaction was significant for *SLC19A1* Pro232Pro ($P = 0.03$) based on the dominant genetic model. The association between lung cancer and the other SNPs evaluated in this study were not modified by smoky coal use. Sex, age, alcohol drinking, family history of cancer, and tobacco smoking did not modify the effect of any of the SNPs examined in this study.

We examined pairwise LD and haplotype block structure for SNPs in the same gene. Haplotypes were estimated for SNPs within the same haplotype block. The results of the haplotype analysis are shown in Table 4. The distribution of haplotypes differed between cases and controls for *CBS* ($P = 0.002$) and *MTHFR* ($P = 0.07$). In addition, three haplotypes in three different genes (*CBS*, *MTHFR*, and *SLC18A1*) were associated with an increased risk of lung cancer compared to the most common haplotype in the gene. In all three genes, the ''at risk'' haplotype contained one of the variants found in this study to be associated with an increased risk of lung cancer.

4. Discussion

Accumulating research results suggest that low levels of folate and elevated homocysteine levels confer increased risk of multiple age-related diseases consistent with its crucial physiological functions [12,29]. Folate is thought to be important in preventing lung cancer because of its roles in DNA synthesis and DNA methylation and some evidence suggests that the protective effect of folate against lung cancer is more evident in heavy smokers [10−12]. The folate pathway may be particularly important for lung cancer risk in the Xuan Wei population because of the extremely high exposure to PAH-rich smoky coal combustion emissions, and proteins involved in the metabolism of folate may modify the protective effects of folate on cancer risk. The role of the proteins coded by these genes is illustrated in Fig. 1.

We studied genetic polymorphisms and haplotypes of folate metabolism genes in a case-control

Table 3 Stratified analysis of *CBS* Ala360Ala and *SLC19A1* Pro232Pro polymorphisms by smoky coal use

	<130 t					>=130 t				
	Cases (%)	Controls (%)	OR[a]	95% CI	*P*-value	Cases (%)	Controls (%)	OR[a]	95% CI	*P*-value
CBS										
Ala360Ala[b]										
TT	10 (20)	24 (36)	Ref.			20 (30)	19 (41)	Ref.		
TC	26 (52)	37 (56)	1.79	0.71—4.52	0.22	36 (54)	23 (50)	1.68	0.71—3.96	0.24
CC	14 (28)	5 (8)	7.86	2.11—29.21	0.002	11 (16)	4 (9)	2.61[c]	0.62—13.02	0.22
TC + CC	40 (80)	42 (64)	2.40	0.99—5.85	0.05	47 (70)	27 (59)	1.82	0.79—4.16	0.16
Trend					0.003					0.13
SLC19A1										
Pro232Pro[b]										
TT	8 (16)	27 (41)	Ref.			18 (26)	11 (24)	Ref.		
TC	26 (52)	23 (35)	4.58	1.66—12.64	0.003	32 (47)	28 (61)	0.61	0.24—1.57	0.31
CC	16 (32)	16 (24)	3.33	1.13—9.82	0.03	18 (26)	7 (15)	1.59	0.48—5.34	0.45
TC + CC	42 (84)	39 (59)	4.01	1.59—10.12	0.003	50 (74)	35 (76)	0.80	0.33—1.94	0.62
Trend					0.02					0.57

[a] Adjusted for age, sex, and current fuel type by unconditional logistic regression.
[b] The interaction with smoky coal use was not significant for *CBS* Ala360Ala ($P = 0.63$), but significant for *SLC19A1* Pro232Pro ($P = 0.03$).
[c] Fisher's exact estimate and test for the parameter without adjustment for other factors.

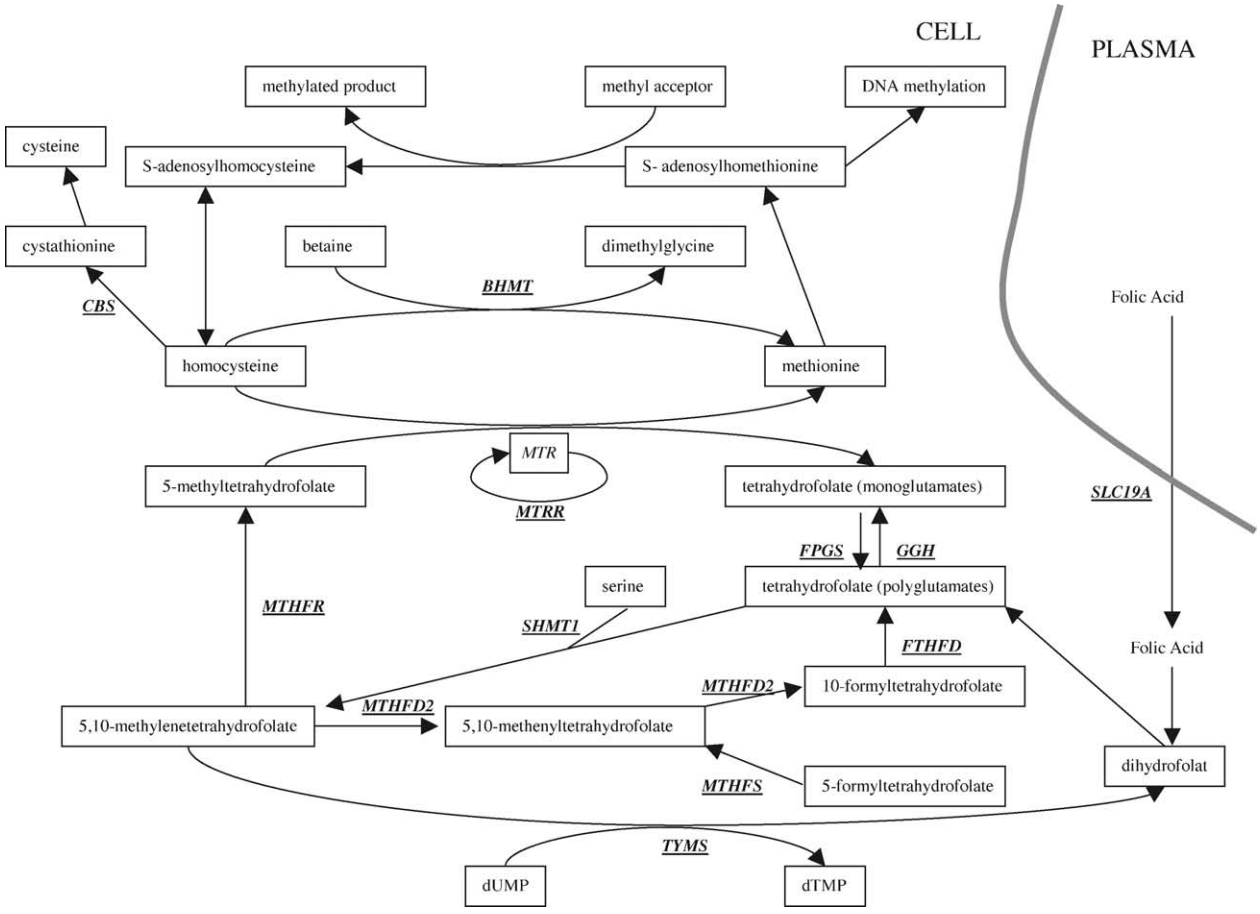

Fig. 1 The role of the proteins coded by studied genes in the folate and methylation metabolism circles.

Table 4 Omnibus test[a] and haplotype analysis of one-carbon genes in Xuan Wei

	Haplotypes	Cases	Controls	OR[b]	95% CI	*P*-value
BHMT	Ex8 +453 A > T—IVS4 +52 C > T					
Hap1	A—C	108	101	Ref.		
Hap2	A—T	100	105	0.89	0.60—1.31	0.54
Hap3	T—C	28	20	1.30	0.69—2.47	0.41
Omnibus test						0.47
CBS	Ala360Ala—Tyr233Tyr					
Hap1	T—C	123	148	Ref.		
Hap2	C—C	106	68	1.89	1.28—2.78	0.001
Hap3	C—T	7	10	0.84	0.31—2.26	0.72
Omnibus test						0.002
FTHFD	Leu395Leu—Asp793Gly					
Hap1	G—A	163	156	Ref.		
Hap2	T—A	44	45	0.92	0.57—1.48	0.73
Hap3	G—G	28	25	1.07	0.60—1.92	0.81
Hap4	T—G	1	0			
Omnibus test						0.86
GGH	-353 G > T—IVS7 −3001 C > T					
Hap1	G—C	120	108	Ref.		
Hap2	G—T	88	93	0.85	0.58—1.26	0.42
Hap3	T—C	28	25	1.02	0.56—1.85	0.96
Omnibus test						0.70
MTHFR	Ala429Glu—Ala222Val					
Hap1	A—C	87	104	Ref.		
Hap2	A—T	103	74	1.67	1.10—2.52	0.02
Hap3	C—C	46	46	1.20	0.73—1.97	0.48
Omnibus test						0.07
MTRR	Leu175Ser—Arg415Cys—His622Tyr					
Hap1	C—C—C	134	110	Ref.		
Hap2	C—C—T	32	39	0.67	0.40—1.15	0.14
Hap3	C—T—C	42	44	0.78	0.48—1.29	0.34
Hap4	T—C—C	28	33	0.70	0.40—1.22	0.21
Omnibus test						0.34
SLC19A1	Pro232Pro—Ex7 −233 G > T					
Hap1	T—G	107	125	Ref.		
Hap2	C—T	116	92	1.48	1.01—2.16	0.04
Hap3	C—G	10	5	2.33	0.77—7.04	0.13
Hap4	T—T	3	2	1.75[c]	0.20—21.29	0.66
Omnibus test						0.15
TYMS	Ex7 +157 T > C—IVS1 −405 C > T					
Hap1	T—C	146	136	Ref.		
Hap2	C—C	10	5	1.88	0.63—5.65	0.26
Hap3	T—T	6	2	2.79[c]	0.49—28.67	0.29
Hap4	C—T	74	83	0.83	0.56—1.23	0.35
Omnibus test						0.32

[a] Exact permutation test.
[b] Adjusted for age, sex and current fuel by unconditional logistic regression.
[c] Fisher's exact estimate and test for the parameter without adjustment for other factors.

study of lung cancer, and found that genetic variants in *CBS*, *MTHFR*, and *SLC19A1* were associated with an increased risk of lung cancer. The protein encoded by *CBS* is involved in the transsulfuration pathway, catalyzing from homocysteine to cystathionine. *CBS* deficiency can cause homocystinuria which affects many organs and tissues. The Ala360Ala polymorphism is common and its frequencies are comparable across different populations [30—32], but to date it has not been studied

with regard to lung cancer risk. Although subjects with the CC genotype have not been observed to have higher homocysteine concentrations in their blood [30—33], the polymorphism may be linked to other functional polymorphisms nearby. It is likely that if the association that we observed between the *CBS* gene and lung cancer is real, it is due to the linkage of the *CBS* Ala360Ala to another functional variant in the region [34,35]; however, a more extensive haplotype analysis would have to be undertaken to confirm this hypothesis.

MTHFR catalyzes the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, a major carbon donor for the remethylation of homocysteine to methionine. Methylenetetrahydrofolate reductase deficiency leads to homocystinuria and decreased levels of methionine, which is a precursor for S-adenosylmethionine, the primary methyl donor for DNA methylation. The 222Val allele in the heterozygous or homozygous state is correlated with decreased enzyme activity and increased thermolability, and individuals homozygous for the mutation have significantly elevated plasma homocysteine levels, particularly in folate-deficient states [18]. In addition, carriers of the 222Val allele display decreased levels of 5-methylcytosine in their genomes [17,36], suggesting that this polymorphism alters DNA methylation patterns, which may be important for the development of lung cancer. Our study found that the 222Val allele was associated with an increased risk of lung cancer, which is consistent with another small study [37]. However, other studies have not observed a significant association between this SNP and lung cancer risk [38—40]. This inconsistency may be due to different background exposure (PAHs) or folate intake.

Folate absorption in the intestine and folate distribution in tissues and cells are important for folate metabolism, apart from a deficient dietary intake. The main uptake pathway of folate compounds into mammalian cells occurs via either the folate receptor or a specialized reduced folate carrier, SLC19A1, which mediates intestinal folate transport and plays a role in maintaining intracellular concentrations of folate. Our study found that homozygous carriers of the C allele at *SCL19A1* had an increased risk of lung cancer compared to homozygous carriers of the T allele. The polymorphism does not lead to an amino acid change in the protein; however, the polymorphism is located in the middle of a conserved domain of this protein [41] and may be in linkage disequilibrium with another functional variant(s) that affects the transport of folic acid into cells resulting in altered intracellular folate pools, and thus decreased biologically utilizable folate levels. Alternatively, the synonymous substitution may affect the mRNA levels of *SCL19A1*, as synonymous polymorphisms have been suggested to alter mRNA stability in other genes [42]. Significant interaction with PAH-rich smoky coal use suggests that folate depletion may abate DNA repair capacity and that DNA damage caused by heavy PAHs exposure is so extensive that the reduced levels of DNA repair due to folate depletion do not come into effect or the cells undergo apoptosis among people with heavy exposure of smoky coal emissions, and as a result, the impact of the unfavorable polymorphism was restricted to people with less exposure to smoky coal only. Detailed studies on the function and mechanism of polymorphisms in this gene and larger case-control studies on lung cancer are warranted.

There are several limitations to this study. We did not collect information on dietary folate intake so we cannot analyze the interaction between folate intake and folate metabolism. Moreover, the sample size of our study is small, we have limited power to detect associations, and there is a possibility that some of our findings are due to chance [43]. However, the FDR-adjusted P values for *CBS* Ala360Ala and *MTHFR* Ala222Val are still statistically significant ($P < 0.05$). A substantially larger case-control study of lung cancer will begin later this year in this region of China, which will provide us an opportunity to replicate and extend these findings.

In summary, we found that SNPs in several genes involved in the metabolism of folate (*CBS*, *MTHFR*, and *SLC19A1*) were associated with an increased risk of lung cancer in Xuan Wei, China, which is a population exposed to high levels of PAHs from smoky coal use and also has a high probability of dietary folate deficiency. These findings support the hypothesis that folate and one-carbon metabolisms play a role in lung carcinogenesis in this population.

Conflict of interest

All the authors have not been paid for the work. This article is not in conflict with financial interests of any organization.

References

[1] Ferlay J, Bray F, Pisani P, Parkin DM. GLOBOCAN 2000: cancer incidence, mortality and prevalence worldwide. Lyon: IARC Press; 2001.

[2] Stewart BW, Kleihues P. World cancer report. Lyon: IARC Press; 2003.

[3] Mumford JL, He XZ, Chapman RS, et al. Lung cancer and indoor air pollution in Xuan Wei, China. Science 1987;235:217—20.
[4] He X, Yang R. Lung cancer and indoor air pollution from coal burning. Yunnan Science and Technology Press; 1994.
[5] Mumford JL, Helmes CT, Lee XM, Seidenberg J, Nesnow S. Mouse skin tumorigenicity studies of indoor coal and wood combustion emissions from homes of residents in Xuan Wei, China with high lung cancer mortality. Carcinogenesis 1990;11:397—403.
[6] Scott JM, Weir DG. Folic acid, homocysteine and one-carbon metabolism: a review of the essential biochemistry. J Cardiovasc Risk 1998;5:223—7.
[7] Speizer FE, Colditz GA, Hunter DJ, Rosner B, Hennekens C. Prospective study of smoking, antioxidant intake, and lung cancer in middle-aged women (USA). Cancer Causes Control 1999;10:475—82.
[8] Jatoi A, Daly BD, Kramer G, Mason JB. Folate status among patients with non-small cell lung cancer: a case-control study. J Surg Oncol 2001;77:247—52.
[9] Kim YI. Folate and carcinogenesis: evidence, mechanisms, and implications. J Nutr Biochem 1999;10:66—88.
[10] Bandera EV, Freudenheim JL, Marshall JR, et al. Diet and alcohol consumption and lung cancer risk in the New York State Cohort (United States). Cancer Causes Control 1997;8:828—40.
[11] Voorrips LE, Goldbohm RA, Brants HA, et al. A prospective cohort study on antioxidant and folate intake and male lung cancer risk. Cancer Epidemiol Biomarkers Prev 2000;9:357—65.
[12] Shen H, Wei Q, Pillow PC, Amos CI, Hong WK, Spitz MR. Dietary folate intake and lung cancer risk in former smokers: a case-control analysis. Cancer Epidemiol Biomarkers Prev 2003;12:980—6.
[13] Blount BC, Mack MM, Wehr CM, et al. Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. Proc Natl Acad Sci USA 1997;94:3290—5.
[14] Herman JG. Epigenetics in lung cancer: focus on progression and early lesions. Chest 2004;125:119S—22S.
[15] Belinsky SA. Gene-promoter hypermethylation as a biomarker in lung cancer. Nat Rev Cancer 2004;4:707—17.
[16] Wei Q, Shen H, Wang LE, et al. Association between low dietary folate intake and suboptimal cellular DNA repair capacity. Cancer Epidemiol Biomarkers Prev 2003;12:963—9.
[17] Paz MF, Avila S, Fraga MF, et al. Germ-line variants in methyl-group metabolism genes and susceptibility to DNA methylation in normal tissues and human primary tumors. Cancer Res 2002;62:4519—24.
[18] Frosst P, Blom HJ, Milos R, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet 1995;10:111—3.
[19] Bagley PJ, Selhub J. A common mutation in the methylenetetrahydrofolate reductase gene is associated with an accumulation of formylated tetrahydrofolates in red blood cells. Proc Natl Acad Sci USA 1998;95:13217—20.
[20] Ronnenberg AG, Goldman MB, Aitken IW, Xu X. Anemia and deficiencies of folate and vitamin B-6 are common and vary with season in Chinese women of childbearing age. J Nutr 2000;130:2703—10.
[21] Hao L, Ma J, Stampfer MJ, et al. Geographical, seasonal and gender differences in folate status among Chinese adults. J Nutr 2003;133:3630—5.
[22] UNICEF. Vitamin & mineral deficiency: a damage assessment report for CHINA, 9-3-2004. United Nations Children's Fund (UNICEF).
[23] Xiao KZ, Zhang ZY, Su YM, et al. Central nervous system congenital malformations, especially neural tube defects in 29 provinces, metropolitan cities and autonomous regions of China: Chinese birth defects monitoring program. Int J Epidemiol 1990;19:978—82.
[24] Pei LJ, Li Z, Li S, et al. The epidemiology of neural tube defects in high-prevalence and low-prevalence areas of China. Zhonghua Liu Xing Bing Xue Za Zhi 2003;24:465—70.
[25] Lan Q, He X, Costa DJ, et al. Indoor coal combustion emissions, GSTM1 and GSTT1 genotypes, and lung cancer risk: a case-control study in Xuan Wei, China. Cancer Epidemiol Biomarkers Prev 2000;9:605—8.
[26] Garcia-Closas M, Egan KM, Abruzzo J, et al. Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. Cancer Epidemiol Biomarkers Prev 2001;10:687—96.
[27] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B 1995;57:289—300.
[28] Lan Q, Chapman RS, Schreinemachers DM, Tian L, He X. Household stove improvement and risk of lung cancer in Xuanwei, China. J Natl Cancer Inst 2002;94:826—35.
[29] Mattson MP, Kruman II, Duan W. Folic acid and homocysteine in age-related disease. Ageing Res Rev 2002;1:95—111.
[30] Aras O, Hanson NQ, Yang F, Tsai MY. Influence of 699C → T and 1080C → T polymorphisms of the cystathionine beta-synthase gene on plasma homocysteine levels. Clin Genet 2000;58:455—9.
[31] Kruger WD, Evans AA, Wang L, et al. Polymorphisms in the CBS gene associated with decreased risk of coronary artery disease and increased responsiveness to total homocysteine lowering by folic acid. Mol Genet Metab 2000;70:53—60.
[32] De S, Dekou V, Nicaud V, et al. Linkage disequilibrium at the cystathionine beta synthase (CBS) locus and the association between genetic variation at the CBS locus and plasma levels of homocysteine. Ann Hum Genet 1998;62:481—90.
[33] Lievers KJ, Kluijtmans LA, Heil SG, et al. Cystathionine beta-synthase polymorphisms and hyperhomocysteinaemia: an association study. Eur J Hum Genet 2003;11:23—9.
[34] Gaustadnes M, Wilcken B, Oliveriusova J, et al. The molecular basis of cystathionine beta-synthase deficiency in Australian patients: genotype-phenotype correlations and response to treatment. Hum Mutat 2002;20:117—26.
[35] Lievers KJ, Kluijtmans LA, Heil SG, et al. A 31 bp VNTR in the cystathionine beta-synthase (CBS) gene is associated with reduced CBS activity and elevated post-load homocysteine levels. Eur J Hum Genet 2001;9:583—9.
[36] Castro R, Rivera I, Ravasco P, et al. 5,10-methylenetetrahydrofolate reductase (MTHFR) 677C → T and 1298A → C mutations are associated with DNA hypomethylation. J Med Genet 2004;41:454—8.
[37] Siemianowicz K, Gminski J, Garczorz W, et al. Methylenetetrahydrofolate reductase gene C677T and A1298C polymorphisms in patients with small cell and non-small cell lung cancer. Oncol Rep 2003;10:1341—4.
[38] Heijmans BT, Boer JM, Suchiman HE, et al. A common variant of the methylenetetrahydrofolate reductase gene (1p36) is associated with an increased risk of cancer. Cancer Res 2003;63:1249—53.
[39] Jeng YL, Wu MH, Huang HB, et al. The methylenetetrahydrofolate reductase 677C → T polymorphism and lung cancer risk in a Chinese population. Anticancer Res 2003;23:5149—52.
[40] Shen H, Spitz MR, Wang LE, Hong WK, Wei Q. Polymorphisms of methylene-tetrahydrofolate reductase and risk of lung

cancer: a case-control study. Cancer Epidemiol Biomarkers Prev 2001;10:397—401.
[41] Tolner B, Roy K, Sirotnak FM. Organization, structure and alternate splicing of the murine RFC-1 gene encoding a folate transporter. Gene 1997;189:1—7.
[42] Duan J, Wainwright MS, Comeron JM, et al. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. Hum Mol Genet 2003;12:205—16.
[43] Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 2004;96:434—42.

Available online at www.sciencedirect.com
SCIENCE DIRECT®